110-IF WORD IS LOWER CASE, THEN ADD WORD TO FST WITH APPROPRIATE EXTENDED GLOSS CODE. IF WORD CONTAINS DECOMPOSABLE CHARACTERS, THEN GENERATE DECOMPOSED VERSION OF WORD AND ADD TO FST WITH APPROPRIATE EXTENDED GLOSS CODE 120-IF WORD IS TITLE CASE OR LOWER CASE, THEN ADD TITLE CASE VERSION OF WORD TO FST WITH APPROPRIATE EXTENDED GLOSS CODE. IF WORD CONTAINS DECOMPOSABLE CHARACTERS, THEN GENERATE DECOMPOSED VERSION OF WORD AND ADD TO FST WITH APPROPRIATE EXTENDED GLOSS CODE 130-IF WORD IS UPPER CASE, LOWER CASE OR TITLE CASE, THEN ADD UPPER CASE VERSION OF WORD TO FST WITH APPROPRIATE EXTENDED GLOSS CODE. IF WORD CONTAINS DECOMPOSABLE CHARACTERS, THEN GENERATE DECOMPOSED VERSION OF WORD AND ADD TO FST WITH APPROPRIATE EXTENDED GLOSS CODE 140-IF WORD IS NEITHER LOWER CASE NOR TITLE CASE NOR UPPER CASE, THEN ADD WORD TO FST WITH APPROPRIATE EXTENDED GLOSS CODE.

FIG. 1

LINGUISTIC DICTIONARY

FINITE STATE TRANSDUCER FIG. 2

EXPLICITLY STORE SUBSTANTIALLY ALL ORTHOGRAPHIC VARIATIONS OF WORDS IN A FINITE STATE TRANSDUCER DATABASE STORE, FOR EACH OF THE ORTHOGRAPHIC VARIATIONS, A CUT AND PASTE CODE EXTENDED BY A GLOSS CODE

STORE, FOR EACH WORD HAVING AN ACCENTED CHARACTER: A WORD HAVING A COMPOSITE FORM OF THE ACCENTED CHARACTER; AND A WORD HAVING AN EXPANDED FORM OF THE ACCENTED CHARACTER.

FIG. 3